Proc. Natl. Acad. Sci. USA
Vol. 74, No. 4, pp. 1458–1462, April 1977
Biochemistry

Secondary structural complementarity between DNA and proteins

(DNA–protein interaction/β-ribbon/DNA minor groove)

GEORGE M. CHURCH, JOEL L. SUSSMAN*, AND SUNG-HOU KIM

Department of Biochemistry, Duke University Medical School, Durham, North Carolina 27710

Communicated by Cyrus Levinthal, January 17, 1977

ABSTRACT **A model for the complex between double-stranded DNA and a β-ribbon (a two-stranded antiparallel β-sheet) of proteins is proposed as one of the possible modes of structural recognition between DNA and proteins. In this model, the contact occurs on the narrow groove of DNA, and the symmetry elements as well as the repeat distances of DNA and the β-ribbon coincide, thus providing favorable complementary contacts.**

The recognition between nucleic acids and proteins is one of the fundamental molecular processes in all living cells. Examples of such interactions are numerous and include complexes between histone and DNA, between repressor and DNA, between restriction endonuclease and DNA, and between transfer RNA and cognate synthetase.

The specific "sequence recognition" between a double-stranded DNA (ds DNA) and a protein probably occurs in two ways: (*i*) the two strands of DNA separate and the base sequence of a single-stranded region is recognized by the protein; and (*ii*) the base-paired double strand itself is recognized. The model proposed here is a general prealignment scheme, a "structural recognition," for the latter type involving a β-ribbon portion of a protein. The term "β-ribbon" is used to describe a segment of a two-stranded antiparallel β-sheet.

Carter and Kraut (1) proposed a model for a double-stranded RNA (ds RNA) and a two-stranded antiparallel β-ribbon, in which the 2′-hydroxyl of the ribose in RNA forms a hydrogen bond to the free carbonyl oxygen of the peptide backbone and the free NH group forms two hydrogen bonds through a water molecule with the ring oxygen and the 2′-hydroxyl oxygen of the next nucleotide on the narrow groove of ds RNA (see Fig. 1 in ref. 1). They also pointed out that, because the narrow groove of ds RNA is so shallow, there is no room in the antiparallel β-ribbon for residues other than those with very small side chain groups. They ruled out the possibility of a similar model for ds DNA because: (*i*) DNA lacks a 2′-hydroxyl group and (*ii*) the minor groove of ds DNA in the B form is narrower than that of ds RNA.

Rigorous model building shows that ds DNA and β-ribbon also have complementary structural features. We propose here a model (two types: P and A) for structural recognition between ds DNA and an antiparallel two-stranded β-structure (β-ribbon) in which not only do all the symmetry elements of DNA and β-ribbon coincide as in the model for ds RNA (1) but also the channel formed between the β-ribbon and the minor groove is large enough to allow the minor groove of ds DNA to be recognized by various amino acid side chains on the inner surface of the antiparallel β-ribbon. This model is different in the hydrogen-bonding scheme from that proposed for ds RNA by Carter and Kraut (1), but it still maintains the same symmetry of ds DNA. A preliminary description of type A (see below) of this model has been published (2).

THE MODEL

Model Building and Refinement. Space-filling (CPK) models and skeletal Watson-Kendrew models were built for ds DNA according to the coordinates of Arnott and Hukins (3) and for an antiparallel β-ribbon similar to those found commonly in protein structures (4). Because ds DNA has two kinds of pseudo 2-fold axes perpendicular to the helix axis for *each* base pair, one in the plane of the base pair and the other between two adjacent base pairs (Fig. 1), we considered only the antiparallel β-ribbon, which also contains two kinds of pseudo 2-fold axes for each two dipeptide pairs (see Fig. 2). As was observed in many protein structures and was pointed out by Chothia (4), β-sheets usually have a right-handed helical twist. In our model building, the twist of the β-ribbon was matched to the helical parameters of the DNA B form. The peptide backbone was then fitted to the DNA while optimizing van der Waals contacts and forming possible hydrogen bonds between DNA and the β-ribbon, at the same time keeping the DNA structure as close to DNA B conformation as possible. There are two ways in which the antiparallel β-ribbon can be fitted into the narrow groove of ds DNA as shown schematically in Fig. 1—the polarities of the adjacent polynucleotide and polypeptide chains can run antiparallel (type A) or they can run parallel to each other (type P).

When two antiparallel oligopeptides form a β-ribbon, every alternating backbone amido NH group is hydrogen bonded to backbone carbonyl oxygens on the opposite strand, leaving one free NH per dipeptide. In both the P and A types, these free NH groups are utilized to form hydrogen bonds to the 3′-oxygens of the polynucleotide backbone, thus utilizing *all* the backbone NH groups to make hydrogen bonds. Likewise, alternating carbonyl oxygens of the backbone are involved in hydrogen bonding to the opposite strand of the antiparallel β-ribbon. The remaining carbonyl oxygens probably form hydrogen bonds with nearby water molecules or amino acid side chains. The unique portion of the complex is composed of one nucleotide and two peptides and is shown in Fig. 2.

The coordinates for the final complex were measured from the skeletal model so built and were "idealized" by a computer program written by Hermans and McQueen (5) and modified by Sussman and Church (unpublished results). This program minimizes the deviation from "standard" bond distances, bond angles, and certain dihedral angles and optimizes van der Waals contacts, by allowing conformational angles of DNA (α, β, γ, δ, ϵ, ζ, χ) and peptide (ϕ, ψ) to vary (Fig. 3) while maintaining helical parameters and symmetry.

The two types (P and A) of refined models are shown in Figs. 4 and 5. The atomic coordinates of the unique portion of the complexes are listed in Table 1 for the two types of models and the DNA B form for comparison. The fitting between DNA and

Abbreviations: ds DNA, double-stranded DNA; ds RNA, double-stranded RNA.

* Present address: Department of Structural Chemistry, Weizmann Institute, Rehovot, Israel.

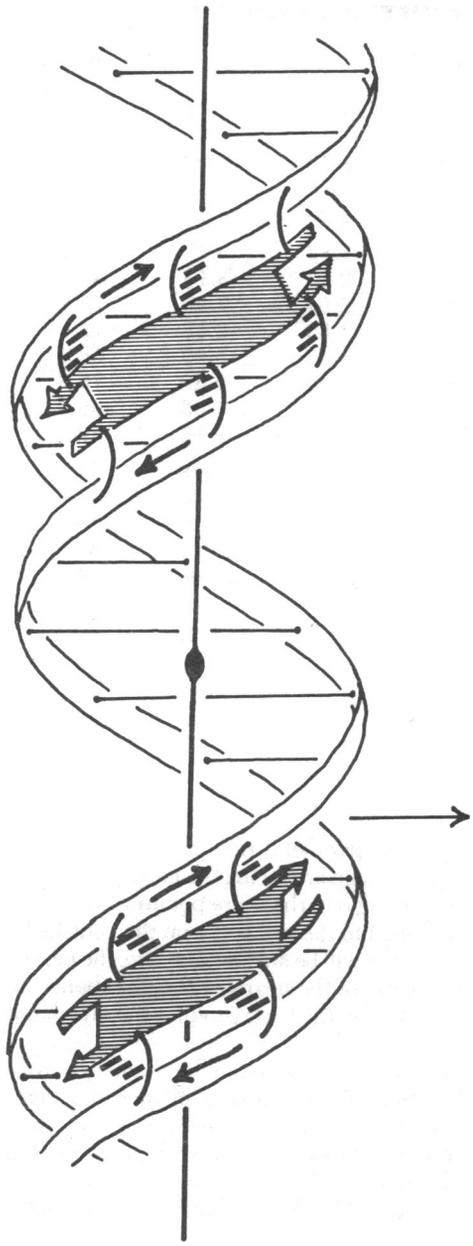

FIG. 1. Schematic drawing of the two types of model complexes: the two DNA backbones are shown as continuous strips wound around the helical axis (vertical line) with the polarity of the polynucleotide (5′ end → 3′ end) indicated with arrows. The two kinds of local 2-fold axes on DNA are shown: one between members of *each* adjacent base pair (an example is indicated with ♦, where the 2-fold axis is perpendicular to the plane of the paper) and the other in the plane of *each* base pair (an example is indicated by →). The β-ribbon formed by a pair of antiparallel, hydrogen-bonded peptides is shaded and the polarity (NH_2 terminus → COOH terminus) is indicated. The three bars between the β-ribbon and DNA represent hydrogen bonds and the curved lines represent charge neutralization between the basic groups on the β-ribbon and the phosphates on DNA. In type A (at top), the polarities of the adjacent DNA and polypeptide backbones are antiparallel; in type P (at bottom), they are parallel.

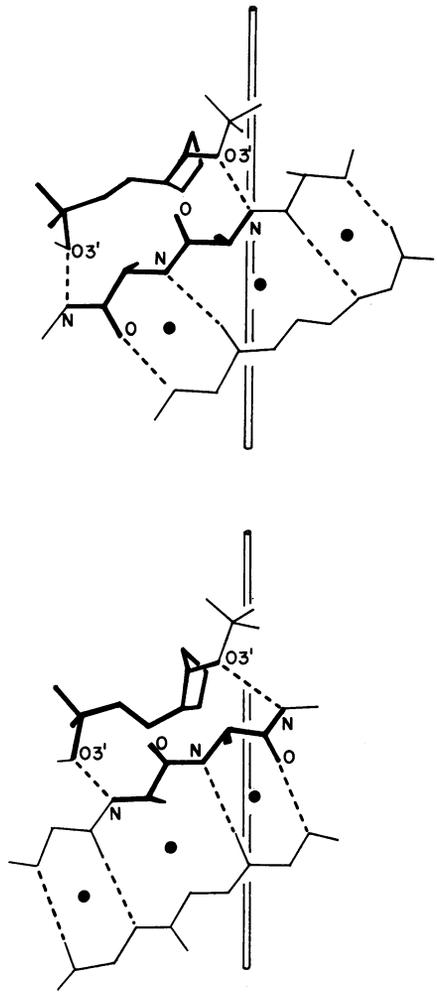

FIG. 2. The asymmetric portions of the DNA–β-ribbon model (type A at top and type P at bottom) are composed of one nucleotide and one dipeptide and are shown in dark lines. Helical axes are shown as vertical rods. The two kinds of local 2-fold axes, originating from the helical axes, intersect on the β-ribbon surface at the point indicated by ●.

the polypeptide backbone is very good. The range of hydrogen bond lengths is between 2.9 and 3.0 Å; there are many good contacts between the peptide and nucleotide backbones, especially among hydrogens attached to saturated carbons such as Cα, C4′, and C5′; and there are no nonbonded contacts shorter than the sum of the respective van der Waals radii [the shortest nonbonded contact is 2.7 Å between the carbonyl oxygen and C4′ in model A, which corresponds to the closest allowed contact distance (6)]. The conformational angles for the model and DNA B form are listed in Table 2. Given the coordinates of this unique portion (Table 1), the rest of the complex can be generated by a simple set of helical parameters: rotation per residue ($\Delta\theta$) = 36°; rise per residue (Δz) = 3.4 Å; two 2-fold axes at (θ = 0°, z = 0 Å) and (θ = 18°, z = 1.7 Å).

Among the two types (P and A) described, the refined stereochemistry of DNA in type P is very close to that of ds DNA B (3), but not in type A. Although ψ, ϕ values of both β-ribbons in the models are within the allowed conformation region in the Ramachandran map (6), the β-ribbon in type A is relatively flat rather than pleated (Fig. 4). The N—H··· O═C—hydrogen bond in type A is bent, as was observed in the β form of polyalanine (7). The major differences between type P and type A are in conformation angles β, γ, δ, and χ of DNA. Although type P has more favorable stereochemistry under the constraint of keeping ds DNA as close to the B form as possible, we can not rule out type A for the following reason: if one allows the helical parameters of ds DNA to vary slightly during the complex formation and the coordinate refinement, the stereochemistry of type A can be improved significantly. Such

Table 1. Cylindrical polar coordinates for the atoms in an asymmetric unit

		Type P			DNA B*			Type A		
Group	Atom	r (Å)	θ (°)	z (Å)	r (Å)	θ (°)	z (Å)	r (Å)	θ (°)	z (Å)
Phosphate	P	9.31	94.7	1.91	8.91	95.2	2.08	10.06	96.1	2.37
	OR	10.59	90.5	1.96	10.20	91.1	1.86	11.42	94.2	2.80
	OL	9.44	101.6	0.98	8.82	103.3	1.29	10.15	101.9	1.29
Ribose	O5′	8.10	88.4	1.47	7.73	88.0	1.83	9.47	88.2	1.72
	C5′	7.76	80.7	2.37	7.70	79.8	2.77	8.13	86.5	1.22
	C4′	7.71	70.7	1.68	7.59	69.9	2.04	7.95	75.8	1.32
	O1′	6.37	66.0	1.60	6.22	66.0	1.83	6.60	72.5	1.66
	C1′	5.86	67.4	0.29	5.86	67.4	0.47	5.99	68.1	0.49
	C3′	8.23	69.9	0.25	8.20	69.9	0.64	8.38	76.6	0.08
	C2′	6.99	72.3	−0.57	7.04	73.2	−0.24	7.03	92.6	−0.62
	O3′	8.90	61.7	0.00	8.75	61.4	0.25	9.18	62.8	0.32
Dipeptide	N	10.25	65.0	3.73				10.98	37.0	1.15
	Cα	10.72	60.5	2.61				10.66	44.7	1.18
	Cβ	12.24	61.0	2.55				9.17	46.2	0.98
	C	10.36	52.4	2.75				10.91	48.9	2.50
	O	10.68	49.2	3.79				11.04	45.6	3.55
	N	9.91	48.7	1.68				10.91	56.1	2.45
	Cα	9.72	40.2	1.65				11.21	60.7	3.62
	Cβ	8.27	37.5	1.90				12.73	61.4	3.71
	C	10.19	36.4	0.35				10.69	68.4	3.54
	O	10.67	40.2	−0.57				10.19	71.0	2.51

Two 2-fold axes are at ($\theta = 0°, z = 0.0$ Å) and ($\theta = 18°, z = 1.7$ Å).
* From ref. 3.

polymorphism of natural DNA has been observed from x-ray fiber diffraction studies (8).

Structual Features of the Model. The ds DNA has two pseudo 2-fold axes per base pair, one on the plane of each base pair and the other between two adjacent base pairs, and both are perpendicular to the helix axis. The antiparallel β-ribbon also has two kinds of pseudo 2-fold axes. In both models, these two pseudo 2-fold axes from ds DNA coincide with those from the antiparallel β-ribbon. In addition, there is also structural complementarity between the DNA backbone and β-ribbon to provide good van der Waals contacts.

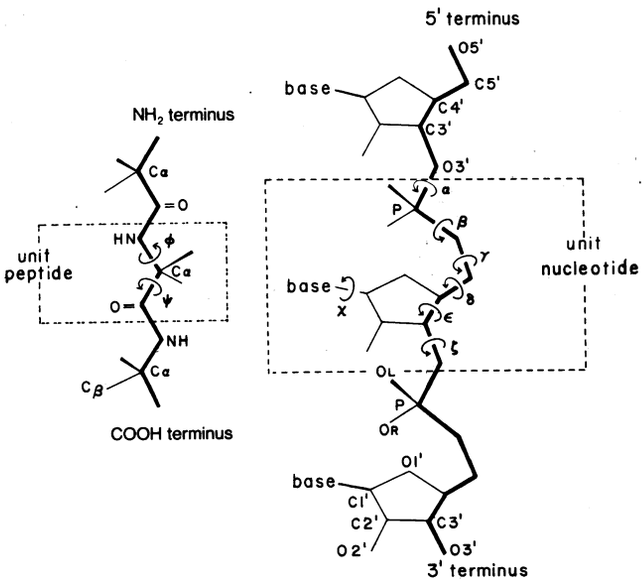

FIG. 3. Nomenclature used in this paper for peptide and nucleotide backbones.

Within the antiparallel β-ribbon, the alternating α carbons of each peptide chain are facing the narrow groove of ds DNA, and the remaining α carbons are pointing outward. Any basic residues attached to the α carbons on the "outside" of the antiparallel β-ribbon will be able to neutralize the negative charges of the phosphates of the nearby DNA backbone. This feature of neutralization by the basic group is shown schematically in Fig. 1.

The tube-like space formed in the narrow groove of ds DNA B by the antiparallel β-ribbon is of such size that almost all amino acid side chains attached to the α carbons on the "inside" surface of the β-ribbon can be accommodated, thus allowing the narrow groove side of base pairs to be specifically recognized by various side chains from the antiparallel β-ribbon.

DISCUSSION

The structural complementarity between the β-ribbon and the narrow groove of ds DNA is likely to provide a strong stabilizing factor in the nonspecific prealignment of a protein and an interacting ds DNA to set the stage for the specific recognition to take place. Two possible examples are the *Escherichia coli lac* repressor–operator complex and the histone–DNA complex.

There is genetic (9, 10) and chemical (11) evidence that at least the first 60 residues of the NH_2 terminus of the *lac* repressor is responsible for DNA binding. The conformation predicted for this region of repressor (12) contains at least two β-regions (residues 4–7 and 17–24) that could form a β-ribbon. Chou *et al.* (12) predicted that this β-ribbon may bind to the *major* groove of DNA. We suggest that the *minor* groove of DNA is structurally more complementary to the β-ribbon. The implication of the *minor* groove as a binding site can also be drawn from the experimental result that actinomycin D competes with *lac* repressor for the *lac* operator site (13) and from a convincing model of actinomycin D binding on the minor

Table 2. Backbone dihedral angles in degrees in an asymmetric unit

	Type P	DNA B*	Type A
Peptide			
ϕ 1	−152.6		−195.0
ψ 1	130.5		178.4
ϕ 2	−144.3		−146.0
ψ 2	181.5		176.3
DNA			
α	−85.0	−95.2	−99.6
β	−47.3	−46.9	60.1
γ	−143.3	−146.0	−208.3
δ	17.3	36.4	−95.6
ϵ	152.7	156.4	149.9
ζ	165.0	155.0	194.2
χ	85.3	82.3	45.4

* From ref. 3.

groove side of the double helix (14) as deduced from x-ray studies of the *a* complex between actinomycin D and guanosine nucleotides (15). Recent chemical modification studies on the complex between the *lac* repressor and random ds DNA also indicate that the narrow groove of ds DNA is better protected by the protein (16).

The secondary structure prediction methods may not be applicable for regions that have a high population of basic residues which interact extensively with DNA. In calf thymus histone H2A, there are two long stretches of sequences containing alternating basic residues (residues 2–17 and 94–105). These alternating residues can all be on the "outside" of the

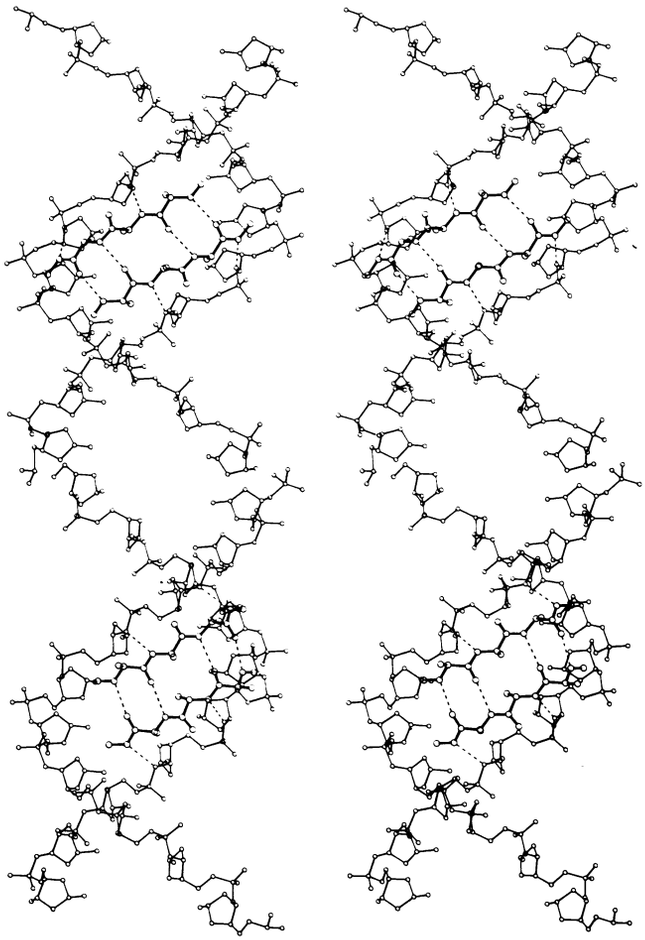

FIG. 4. Stereo view of the DNA–β-ribbon atomic model (type A at top; type P at bottom). For clarity, base pairs in DNA and side chains of the polypeptide are not shown. The peptide backbones in the β-ribbon are shown in thicker lines and the hydrogen bonds are shown in broken lines.

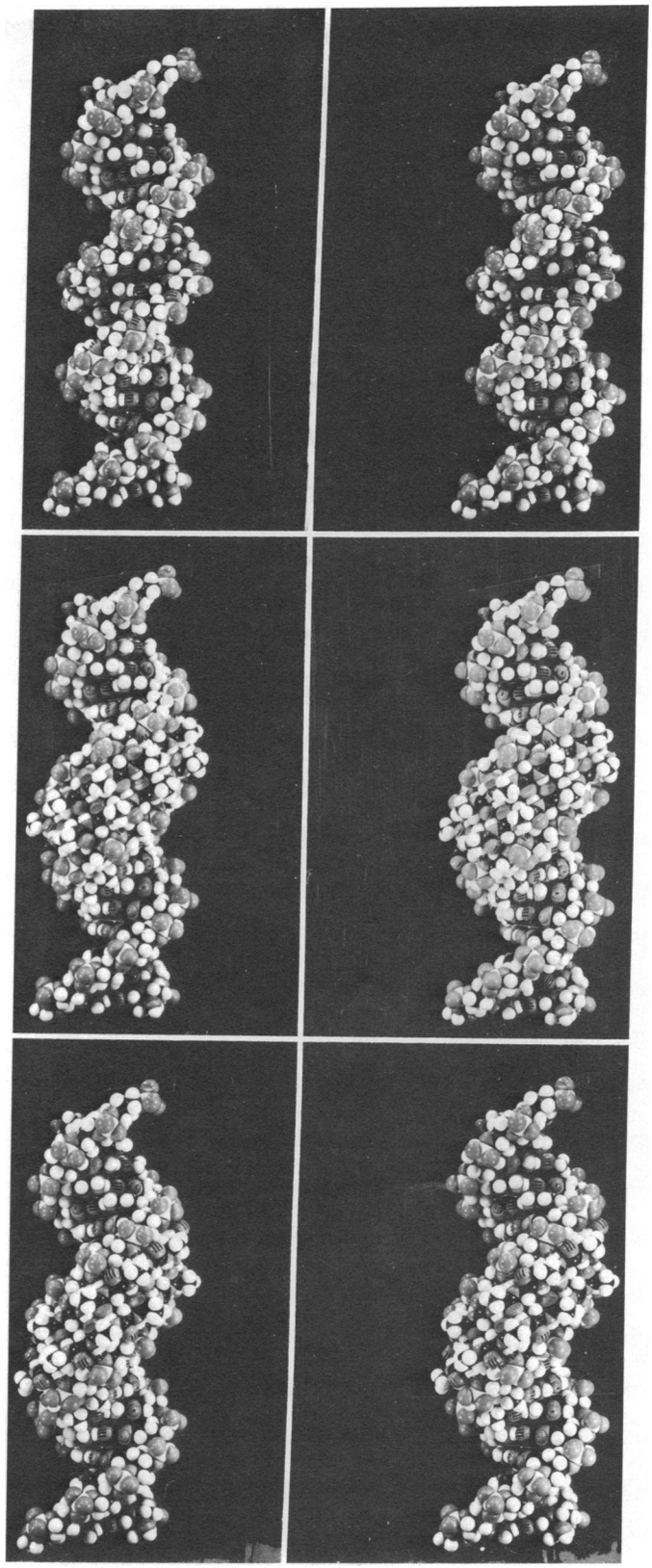

FIG. 5. Stereo views of space-filling models for DNA B (*top*), model complex type A (*middle*), and type P (*bottom*).

β-ribbon and could neutralize the negative charge of phosphate groups of DNA as shown in Fig. 1.

The model proposed here is a plausible way in which an antiparallel β-ribbon of a *portion* of a protein can interact with ds DNA in a nonspecific way by coinciding pseudo 2-fold axes of both secondary structures and by forming hydrogen bonds between them. The model also shows sufficient room to accommodate various side chains on the "inside" of the β-ribbon to interact with base pairs on the narrow groove of ds DNA. Among the amino acid residues, all except those of proline and tryptophan can be accommodated in the space between the β-ribbon and the narrow groove of the DNA in both models. The side chains of the following residues can form one or more hydrogen bonds with base(s) depending on the base sequence in the immediate neighborhood of the side chains: Arg, Asn, Asp, Gln, Glu, His, Lys, and Tyr. Any additional interaction on the wide groove and backbone of ds DNA will further stabilize the complex and increase the specificity of the recognition between ds DNA and the protein.

This research was supported by grants from the U.S. Public Health Service (CA 15802 from the National Cancer Institute) and the National Science Foundation (GB 40814). J.L.S. is a fellow of the Arthritis Foundation, and G.M.C. was a National Science Foundation Predoctoral Fellow.

1. Carter, C. & Kraut, J. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 283–287.
2. Kim, S. H., Sussman, J. L. & Church, G. M. (1975) in *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions*, eds. Sundaralingam, M. & Rao, S. T. (University Park Press, Baltimore, MD), pp. 571–575.
3. Arnott, S. & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1504–1509.
4. Chothia, C. (1973) *J. Mol. Biol.* **75**, 295–302.
5. Hermans, J. & McQueen, J. L. (1974) *Acta Crystallogr. Sect. A* **30**, 730–739.
6. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437.
7. Arnott, S. & Dover, S. D. (1967) *J. Mol. Biol.* **30**, 209–212.
8. Bram, S. & Tougard, P. (1972) *Nature New Biol.* **239**, 128–131.
9. Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Mueller-Hill, B., Pfahl, M. & Schmitz, A. (1972) *Nature* **237**, 322–327.
10. Weber, K., Platt, T., Ganem, D. & Miller, J. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3624–3628.
11. Platt, T., Files, J. G. & Weber, K. (1973) *J. Biol. Chem.* **248**, 110–121.
12. Chou, P. Y., Adler, A. J. & Fasman, G. D. (1975) *J. Mol. Biol.* **96**, 29–45.
13. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67–83.
14. Sobell, H. M. & Jain, S. C. (1972) *J. Mol. Biol.* **68**, 21–34.
15. Jain, S. C. & Sobell, H. M. (1972) *J. Mol. Biol.* **68**, 1–20.
16. Kolchinsky, A. M., Mirzabekov, A. D., Gilbert, W. & Li, L. (1976) *Nucleic Acid Res.* **3**, 11–18.